METEOROLOGICAL APPLICATIONS
Meteorol. Appl. **21**: 708–716 (2014)
Published online 2 July 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/met.1396

On the application of hierarchical cluster analysis for synthesizing low-level wind fields obtained with a mesoscale boundary layer model

Gustavo Ratto,[a,b]* Guillermo J. Berri[c,d] and Ricardo Maronna[e]

[a] *CIOp (Centro de Investigaciones Opticas), La Plata, Argentina*
[b] *Facultad de Ingeniería, Universidad Nacional de La Plata, La Plata, Argentina*
[c] *SMN (Servicio Meteorológico Nacional), Buenos Aires, Argentina*
[d] *CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), Buenos Aires, Argentina*
[e] *Facultad de Ciencias Exactas, Departamento de Matemáticas, Universidad Nacional de La Plata, La Plata, Argentina*

ABSTRACT: Hierarchical clustering is applied to a boundary layer model output that describes the low-level wind field over the La Plata River region of South America. The model output consists of 180 17-dimensional vectors *per* season that include wind direction frequencies, calms and mean wind speeds *per* wind sector. The cluster approach is intended to assist the discussion of meteorological phenomena, and is also employed to define regionality. Results show that the 180 original vectors can be well represented by a small number of vectors, and the 18, 12 and 6 group cluster solutions share a similar layout. However, the 12 and the 6 group clusters seem both appropriate solutions when a threshold of 10% in wind direction frequency, including calms, is taken as a reference in order to decide significant differences between groups. All solutions show more groups along the northeastern than along the southwestern river shore, evidencing a complex sea-land breeze circulation pattern. The analysis of the observations at nine weather stations supports the findings of the cluster analysis conducted with the model outputs. The advantage of the hierarchical cluster analysis in synthesizing information becomes clearly evident when compared to the traditional method of visual inspection. Besides, the actual distribution of weather stations in the region is not very far from the regionality that suggests the obtained cluster distribution. However, in order to match the latter, more observing points would be needed particularly over the river and towards the northeastern shore.

KEY WORDS La Plata River region; model outputs; multivariate analysis; surface winds; weather station siting

Received 18 May 2012; Revised 24 January 2013; Accepted 7 March 2013

1. Introduction

Berri *et al.* (2010) present a mesoscale boundary-layer model, BLM, and a simplified methodology used to simulate the high-horizontal-resolution low-level wind field 'climatology' over the La Plata River region of South America. The model, forced with local weather observations, is able to reproduce the regional wind fields with a reduced number of daily forecasts that, according to the authors, are reasonably accurate.

Cluster analysis is an exploratory technique that provides an objective way of grouping individuals (vectors) in such a way that helps in summarizing information. This technique has been employed successfully for the examination of regions of climate variability (Wolter, 1987; Fovell and Fovell, 1993; Gong and Richman, 1995; Unal *et al.*, 2003). In previous reports (Ratto *et al.*, 2010a, 2010b) hierarchical cluster analysis has been applied to summarize information regarding hourly wind roses.

In the present work hierarchical clustering is applied to the BLM output over the inner rectangle of Figure 1, indicated as La Plata River region, aimed at defining areas of spatial homogeneity in the wind fields. The model output consists of a set of 17-dimension vectors that include 8 sector wind direction frequencies, frequency of calms, and mean wind speed of each wind sector. The use of hierarchical clustering in this work follows two purposes. The first one is to synthesize wind field information employing an objective tool in order to assist the discussion and interpretation of meteorological phenomena in the area under study. The second purpose of this paper is to discuss meteorological regionality by identifying homogeneous sub areas of similar wind field characteristics. In this sense, the cluster analysis is intended as a 'design tool' that deals with the possibility of suggesting an optimal number of representative weather stations that should operate in the region, one in each of the obtained sub areas. The cluster analysis is strictly performed from a statistical point of view, but the resulting regionality is interpreted from a meteorological point of view.

2. Climatological characteristics of the region

The La Plata River is considerably wide (between 40 and 200 km), so that the significant land-river surface temperature contrast establishes a low-level circulation, with sea-land breeze characteristics. The daily cycle of the differential heating gives rise to significant changes of the predominant wind direction across the region throughout the day, as can be appreciated in Figure 2 (the inner rectangle indicates the region in which the cluster analysis is performed). This figure shows the observed mean wind direction frequencies at the weather stations of the

* Correspondence : G. Ratto, CIOp (Centro de Investigaciones Opticas), CC 124, 1900 La Plata, Provincia de Buenos Aires, Argentina
E-mail: gustavratto@gmail.com

Correction added on 1 November 2013 after original online publication: in the author affiliations the address for SMN and CONICET has been corrected to 'Buenos Aires, Argentina'.

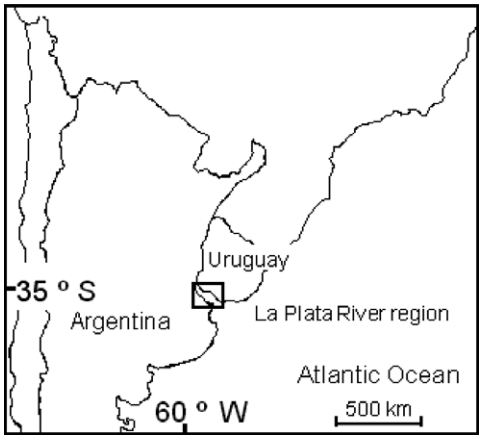

Figure 1. Location of La Plata river region in South America. The inner rectangle indicates the mesoscale model sub domain of the cluster analysis.

region during the period 1994–2008. At 0600 local time (local time is UTC- 3 h) (Figure 2(a)), early morning, the weather stations around the coast of Uruguay show predominantly north and northeast winds and, in general, only minor contributions of other wind directions. Over Argentina the wind directions are more evenly distributed among the different sectors, although north and south show greater frequencies. At 1800 local time (Figure 2(b)), the weather stations near the northern shore of the river show predominantly east, southeast and south winds, while the east winds are clearly dominant in the weather stations over Argentina. Over Uruguay, the early morning offshore winds acquire a strong inland component in the afternoon. Over Argentina the inland wind component is very strong in the afternoon, in particular towards the river springs. Throughout the day the weather stations of the region display a significant change of the predominant wind directions of more than one quadrant.

3. Data employed

The data consist of a grid of 180 points (18 in the west–east direction and 10 in the south–north direction), over the inner rectangle depicted in Figure 2. At each point, the wind climatology at 10 m is calculated with the BLM. The model output consists of 17-dimension vectors that includes 8-sector wind direction frequencies (the first eight variables), frequency of calms (the ninth variable), and mean wind speed of each wind sector (the last eight variables). For simplicity, the first nine variables of the vector are referred to as the wind direction frequency rose (a wind direction rose that includes calms), and the last eight variables of the vector as the mean wind speed rose.

However, a model never calculates a zero wind speed so that it is necessary to adopt a wind speed threshold below which the model output can be considered as calm condition. Berri *et al.* (2010) run different tests with the model and determined that 1 m s^{-1} is the wind speed value below which the resulting percentage of calm winds matches, in the average, the observations over the region, so that in the present study the same value is adopted as the wind speed threshold for calms.

The BLM wind climatology is the ensemble result of a series of daily forecasts obtained by forcing the model with local observations. Each ensemble member produces a daily forecast that participates in the definition of the wind climatology with a probability calculated with the local observations. The upper boundary condition is taken from the only local radiosonde observation available, one a day in the region (EZE in Figure 2). The lower boundary condition consists of a surface heating function calculated with the temperature observations of the surface weather stations of the region. The reader is referred to Berri *et al.* (2010) for details of the ensemble method used for calculating the wind climatology, as well as the BLM formulation. For the purpose of a brief description, it can be said that the BLM has been specifically developed for modelling the low-level circulation over the La Plata River region. The model is dry and hydrostatic, and is based on conservation equations of momentum; mass and heat, with a first-order turbulence closure. The domain of the BLM runs is centred over the La Plata River region and consists of 79 points in the x direction and 58 points in the y direction, with a spatial resolution of 5 km. The horizontal domain is 390 km in longitude and 285 km in latitude, and contains the inner rectangle of Figure 1 in which the cluster analysis is performed. Although the model output is at 5 km horizontal resolution, for simplicity only 1 out of 3 grid points is used to define the 180 point grid of the cluster analysis. The vertical domain has 12 levels up to the material

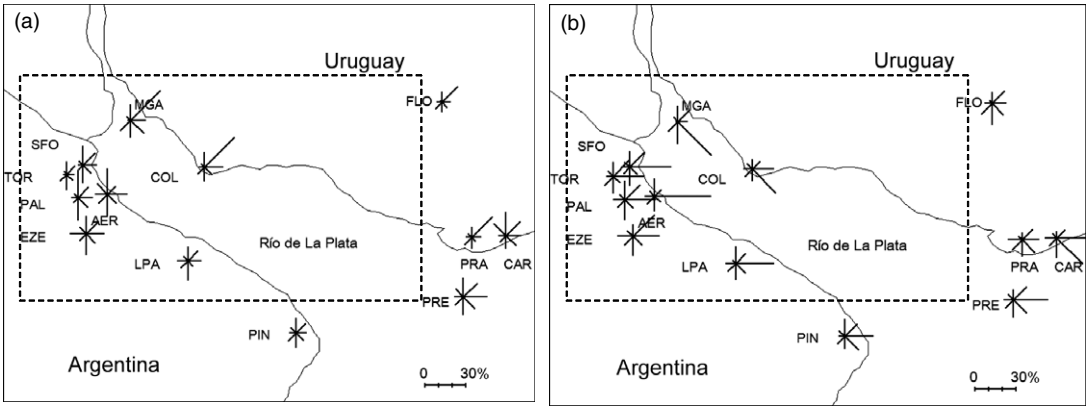

Figure 2. Observed 1994–2008 mean wind direction frequencies, in percentage, at (a) 0600 local time and (b) 1800 local time. The inner rectangle indicates the region in which the cluster analysis is performed. The weather stations, in alphabetical order are: Aeroparque (AER), Carrasco (CAR), Colonia (COL), Ezeiza (EZE), Florida (FLO), La Plata Aero (LPA), Martín García (MGA), El Palomar (PAL), Punta Indio (PIN), Prado (PRA), Pontón Recalada (PRE), San Fernando (SFO) and Don Torcuato (TOR). North is upward in the figure.

top at 2000 m, distributed according to a log–linear spacing, and the 10 m level is the one used.

Observational data from the weather stations in the inner rectangle of Figure 2 are employed as reference in order to be compared with model outputs.

4. Methods

Details on hierarchical clustering are given by Romesburg (2004) and applications to atmospheric sciences can be found in Wilks (2006). The procedure to carry out the cluster analysis in this paper has been outlined in detail in Ratto *et al.* (2010b). Nevertheless, a brief description of the methodology and the process is provided. The clustering process has been carried out with Statistica 8.0 software. Among the different approaches to clustering the authors chose agglomerative hierarchical clustering, because it yields a coherent range of possible classifications and therefore allows the user to choose the most adequate number of groups. Other methods, such as the popular K-means, require some *a priori* knowledge of the number of groups, which is not the present case, and the clusterings corresponding to different numbers of groups are not nested, which makes it more difficult to choose the adequate one.

As the 17-dimensional vectors contain variables of different nature, i.e. wind direction frequencies and calms (%) and mean wind speeds ($m\,s^{-1}$), each variable was normalized with the mean and the standard deviation (Wilks, 2006). The dendrograms (e.g. that of Figure 3) for the seasons have been built using the squared Euclidean distance as a dissimilarity measure (two objects will be consider similar to each other as far as they have a smaller distance). This metric has been largely used in atmospheric sciences (Unal *et al.*, 2003). The 'average linkage between groups' has been chosen as the clustering method because it is less sensitive to outliers (Figueras, 2001), also such a method has been proved to provide the most realistic results in climatological research (Kalstein *et al.*, 1987). The normalized set of 180 vectors of 17 dimensions was clusterized. Firstly, the distance between all original vectors (output of the model) was estimated. The two closest ones were then grouped and the distances between all the vectors are estimated. The process continues by repeating the steps given above until all the original vectors conform one group. The criterion of selecting an appropriate number of groups deals with the compromise between specificity (given by a large number of clusters) and generality (given by a strong reduction of the original cases). In this sense, the optimum number of clusters plays an important role. Three cut- off distances (i.e. three possible cluster solutions discussed in Section 5) that divide the whole area covered by the model into 18, 12 and 6 sub areas have been adopted, in order to discuss solutions of different level of detail (exploratory approach). The results are interpreted and discussed in the context of the climatological aspects of the regional low level wind fields. Each sub area

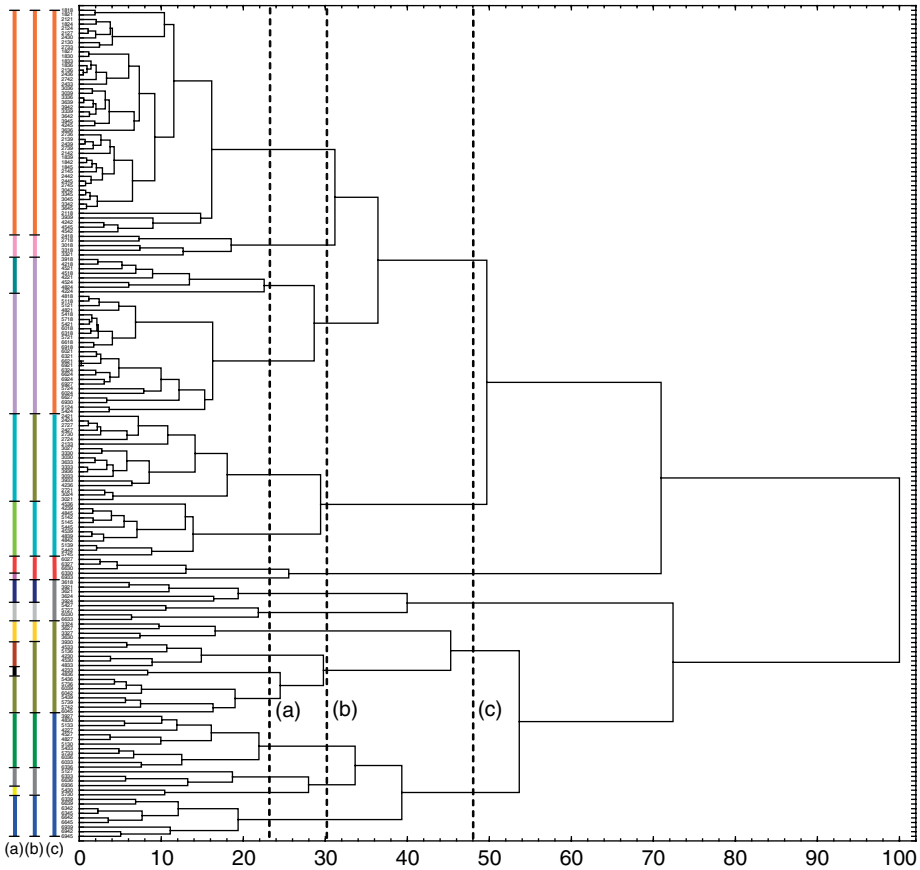

Figure 3. Dendrogram for summer. The right-hand column of the *Y*-axis contains the identification of each of the 180 original vectors, given by a four-digit number (small size) that corresponds to the coordinates of the pixels in Figure 4. The horizontal scale is for the squared Euclidean distance, that appears rescaled with respect to the maximum distance. The three cut-off distances selected for the analysis (23, 30 and 48) are shown with dashed vertical lines. For each of the three respective classifications (cases (a), (b) and (c) of Figure 4), the cluster corresponding to each vector is indicated by a colour line on the left-hand side of *Y*-axis.

is represented by an average vector. Note that the dendrogram (Figure 3) shows a rescaled distance that makes it possible to compare dendrograms corresponding to other seasons.

The well known 'sum for the absolute values of the differences', *SAD*, is a metric that quantifies differences between vectors by addressing the 'distance' between them. As the Euclidean distance, *SAD* is often employed to estimate how different may be one vector from another. In this paper, *SAD*, see Equation (1), is mainly used to assess differences between wind direction frequency roses (as defined in Section 3) and between mean wind speed roses:

$$SAD = \sum_{i=1}^{n} |x_i - y_i| \quad (1)$$

where i is the variable (direction, calm or wind speed) of the pattern (wind rose) involved, n is between 1 and 9 for wind direction frequency roses and between 1 and 8 for mean wind speed roses, x_i is the percent wind direction frequency (including calms) or the mean wind speed ($m\,s^{-1}$) of the pattern **X** in the direction i, y_i is the percent wind direction frequency (including calms) or the mean wind speed ($m\,s^{-1}$) of the pattern **Y** in the direction i.

Applied to quantify differences between observed and predicted vectors, *SAD* provides a degree of 'error' in model outputs (Section 5.1).

Applied to the discussion of differences within model outputs (wind direction frequencies including calms) a $SAD \geq 10\%$ is considered as a limit to decide whenever the difference between the two vectors involved is large enough to be meteorologically meaningful (Sections 5.3–5.6). Taking into account that each wind rose satisfies that the sum of all wind direction frequencies and calms adds 100% and Equation (1), a single proof shows that $|x_i - y_i| \leq \frac{SAD}{2}$ for any i. That is to say, a *SAD* threshold of 10% implies a maximum difference of 5% in one variable of the vector which can be considered negligible from a climatological point of view. Additionally, differences between mean wind speeds were not considered important. This is due to the fact that the differences in mean wind speed among the grid points of the model are considerably smaller than those observed for the wind direction frequencies. Table 1 shows, as an example, two vectors **X** and **Y**. Computing *SAD* with Equation (1) for the vectors of Table 1 gives a value of 3.1% for frequencies and a $0.3\,m\,s^{-1}$ for wind speeds. Since the *SAD* value is, in this case, smaller than the adopted 10% threshold, the difference between vectors is considered irrelevant.

5. Results and discussion

5.1. Comparison between model outputs and observed data

So as to provide a basis to the subsequent cluster analysis, a comparison between average observations and model outputs is carried out.

Table 2. *SAD* values computed for the weather stations within the region under analysis (inner rectangle of Figure 2, see text for details about PRE).

Site	SAD (%)	SAD ($m\,s^{-1}$)
AER	30.1	12.0
COL	20.4	18.0
EZE	25.1	7.5
LPA	28.5	9.2
MGA	30.5	2.9
PAL	29.9	6.2
PRE	9.2	28.8
SFO	20.1	10.0
TOR	24.0	8.3

SAD (%) expresses the sum of the absolute values of differences between observed and modelled wind direction frequency roses (including calms) at the closest grid point to every weather station. *SAD* ($m\,s^{-1}$) is analogous but for average wind speeds.

On one hand the wind direction frequency roses (including calms) and mean wind speed roses observed during the period 1994–2008 at the weather stations within the area of the cluster analysis (inner rectangle of Figure 2) are available. PRE was included in this comparison (slightly out of the inner rectangle of the figure) because of its strategic location for the study. PRE is located in a wide homogeneous water surface with small horizontal gradients so that its observations are representative of the conditions at the river mouth.

The corresponding wind roses obtained with the model at the closest grid point to every weather station are also available. The maximum distance between a model grid point and a weather station is 3.54 km (the model resolution is 5 km).

The *SAD* values computed with these two vectors (observed and modelled) are shown in Table 2. As a whole, these results are in accordance with the root mean square values of relative errors discussed in Berri *et al.* (2010), for the climatological low-level wind field obtained with the BLM model for the period 1959–1984 over the same region. The authors explain that large errors in some locations are due to their proximity to the coasts, for example AER and MGA, so that the 5 km horizontal resolution of the model becomes a limiting factor. In the particular case of PRE, with a large wind speed SAD, the authors also found large wind speed model errors which the attribute to the fact that the instrument is on ship at 22 m, instead of the standard height of 10 m.

5.2. Dendrogram analysis

Figure 3 (dendrogram for summer) is an example of the clustering process output carried out for the four seasons of the year. The X axis shows a rescaled distance which is appropriate when comparisons with the dendrograms for other seasons are needed. The dendrogram reflects the clustering of the 180 original vectors, showing the progress and the final result in

Table 1. Percent wind direction frequencies including calms and average wind speeds ($m\,s^{-1}$) are the variables of vectors **X** and **Y**.

	Wind direction frequencies (%)									Wind direction mean speed ($m\,s^{-1}$)							
	N	NE	E	SE	S	SW	W	NW	Calms	N	NE	E	SE	S	SW	W	NW
X	11.1	17.3	11.1	15.4	11.2	5.7	4.4	7.3	16.5	3.0	3.3	3.6	3.4	3.0	3.0	2.7	2.7
Y	11.3	17.8	10.5	15.2	11.8	5.6	4.1	6.9	16.8	3.0	3.3	3.6	3.5	3.1	3.1	2.7	2.7

Note that wind direction frequencies and calms add up 100% for each vector.

which the original vectors are forming only one group. The summer is selected (the rest of the seasons are not shown due to space constraints) because it is the season that shows highest variability among the original individual vectors, so that it is the best suited example to discuss the usefulness of the proposed method. The variance for wind speeds is very similar throughout the seasons, but for wind directions summer presents the greatest variation (30.9) compared to the rest of the seasons (16.9 for spring, 14.8 for autumn and 9.5 for winter). When comparing seasonal dendrograms, colder seasons (autumn and winter) show that more groups are agglomerated in shorter relative distances compared to warmer seasons (spring and summer). For example, for a relative distance of 40 (see *X* axis in Figure 3), summer shows 8 groups while winter (not shown) only 6.

Besides, the comparison between the seasonal dendrograms allowed us to infer that stable cut-off relative distances (distances in which the groups appear better separated) are, in general, greater than 30. This implies that the region under study can be well represented with a small number of vector groups (or sub areas).

According to the cluster analysis, a first conclusion that can be drawn is that the low-level climatological wind field of the region could be reasonably well represented by a small number of wind roses.

Three cut-off distances of 48, 30 and 23 (see the dash vertical lines in Figure 3) were taken to explore the data structure, giving 6, 12 and 18 groups of vectors whose corresponding sub areas are shown in Figure 4. The goal is to analyse which of the three clustering outputs summarizes better the information provided by the model.

Each coloured rectangle of Figure 4 is represented by a wind direction frequency rose (including calms) and a mean wind speed rose in Figure 5. Each wind rose of this figure, identified with a specific colour, is the average of the original model output.

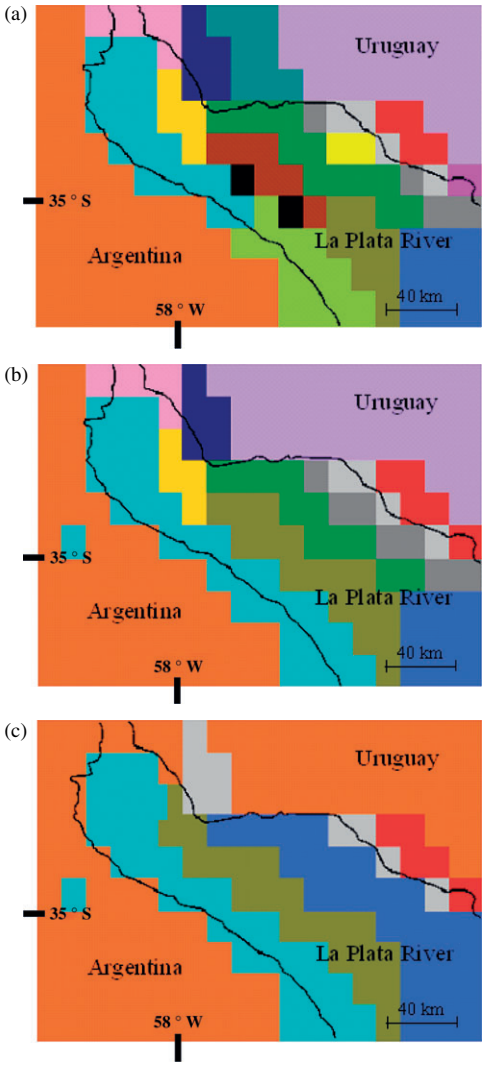

Figure 4. The inner rectangle of Figure 1 is shown divided into (a) 18 sub areas (b) 12 sub areas and (c) 6 sub areas according to the 3 cluster solutions for summer, adopted as the example to be shown. Each sub area identified with a particular colour contains a particular number of pixels (each pixel is approximately 15×15 km a side), according to the corresponding cluster solution.

5.3. Similarities between clusters for the 18-group cluster solution

By visual inspection, it is possible to detect strong similarities between, for example, cyan and apple green clusters or between yellow and dark grey clusters. To evaluate differences within the 18-cluster the *SAD* is calculated, Equation (1), obtaining a value of less than 10% for the wind direction frequency roses between cyan and apple green clusters and between yellow and dark grey clusters (see also Figure 4(a)), and a negligible *SAD* value for the wind speeds roses. The rest of the *SAD* values between clusters for the wind direction frequencies (including calms) were all above 10%. The maximum *SAD* for mean wind speed roses between groups of the 18-group cluster solution was 3.6 m s^{-1}.

5.4. Similarities between 18-group and 12-group cluster solutions

In the same way, the differences between the 18 and the 12-group cluster solutions (Figure 5(a) and (b)) are analysed. A visual inspection reveals a strong similarity, for example, between the violet clusters of the 18 and the 12-group cluster solutions.

The *SAD* value between the violet, red and dark yellow clusters of the 12 and 18-group cluster solutions, as well as between dark grey and yellow clusters of the 12 and 18-group cluster solutions, respectively, are all below 10% for wind direction and below 5 m s^{-1} for wind speeds. The comparison between the rest of the clusters of these two clustering solutions gives *SAD* values above 10% for wind direction.

These results suggest that the 18-group cluster solution may be at first glance somewhat redundant. In order to validate this idea the same analysis is used to compare the 6 and the 12-group cluster solutions and within the 6-group cluster solution itself.

5.5. Similarities between 12-group and 6-group cluster solutions

Only two of the clusters of the 6 and 12-group cluster solutions have differences below 10% for wind direction, while wind speeds do not differ significantly. The comparison between the rest of the vectors of these two clustering solutions have *SAD* values growing rapidly above 10% for wind direction.

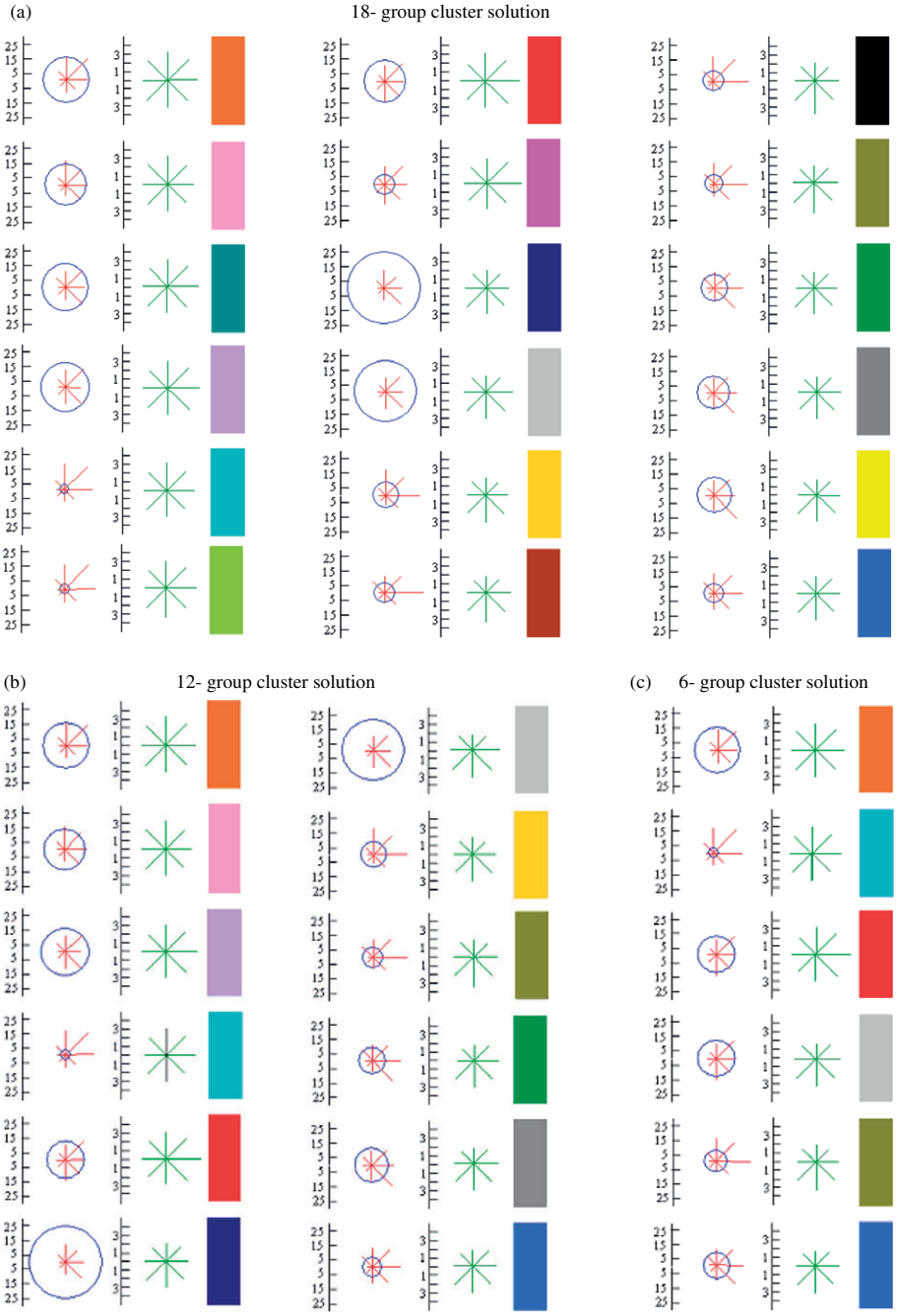

Figure 5. Resultant wind roses for the three cluster solutions (wind direction frequencies including calms are given in percentage while mean wind speeds by sector are given in m s^{-1}). (a) Wind direction frequency roses (red lines) including calms (blue circles), both expressed as percentage of occurrences. *Y*-axis represents the percent frequencies for the corresponding wind direction, including calms. Right side: Wind speed roses (green lines) expressed in m s^{-1}. *Y*-axis represents the average wind speed for the corresponding wind direction. Each wind rose is the result of averaging the corresponding data in accordance with the formed clusters for the three summer solutions. (a) Corresponds to the 18-group cluster solution of Figure 4(a) and the relative distance of 23 in Figure 3(a). (b) Corresponds to the 12-group cluster solution of Figure 4(b) and the relative distance of 30 in Figure 3(b). (c) Corresponds to the 6-group cluster solution of Figure 4(c) and the relative distance of 48 in Figure 3(c). Each pair of wind roses, identified with a unique colour in this figure, covers the sub area of the same colour in Figure 4.

5.6. Similarities between clusters for the 6-group cluster solution

The minimum *SAD* within vectors for the 6 cluster solutions were 17.2% for wind direction frequency roses while 1.9 m s^{-1} for wind speed roses.

At this point it is possible to conclude that the 12 and the 6-group clusters seem both appropriate solutions when a threshold of 10% in wind direction frequency including calms is taken as a reference in order to decide significant differences between groups.

5.7. Comparison among observations in sub areas

In order to support the results of the cluster analysis obtained from the BLM outputs, we calculate the *SAD* among observed

Table 3. *SAD* values computed for the weather stations located in continental Argentina, within the region under analysis (inner rectangle of Figure 2).

Site	AER	EZE	LPA	PAL	TOR	SFO
AER		36.0	25.8	20.9	22.1	20.5
EZE			15.5	17.9	17.2	17.3
LPA				11.3	8.6	14.4
PAL					13.4	14.6
TOR						10.3

SAD (%) expresses the sum of the absolute values of differences between observed wind direction frequency roses (including calms).

wind direction roses at the weather stations within the study region, including PRE (please see the considerations discussed in Section 5.1 regarding this inclusion). Table 3 shows the *SAD* values among weather stations in continental Argentina. The averaged *SAD* value among all the stations is 17.7%. The first row of Table 3 shows that the larger values correspond to AER, with an averaged *SAD* of 25.1% with the rest of the stations. In particular, the *SAD* between AER and EZE is 36.0%.

Excluding AER, the averaged *SAD* of Table 3 is 14.1%, quite close to the 10% threshold adopted for the cluster analysis. Except for AER, the rest of the weather stations in continental Argentina display similarity of wind direction roses, which agrees with the results of the cluster analysis that finds this region basically comprising one sub area.

AER is located only a few hundred metres inland from the river shore, so close to the river-land surface thermal contrast boundary that its behaviour may reflect singularities of the interaction between continental and maritime meteorological regimes.

Table 4 shows the *SAD* values among AER, EZE and the other weather stations in the river and Uruguay. The averaged *SAD* value among all these stations is 38.6%, more than twice that of 17.1% corresponding to the group of 6 stations in continental Argentina. It must be pointed out that all these stations are located in different sub areas, or very close to their borderlines. The averaged *SAD* among weather stations outside continental Argentina, i.e. PRE, COL and MGA, all located in different sub areas, is 36.6%, The averaged *SAD* PRE-COL-MGA plus EZE is 32.4%, while the averaged *SAD* PRE-COL-MGA plus AER climbs to 43.9%. Either one of these two values is more than double the averaged *SAD* value of 14.1% among weather stations inland continental Argentina.

The analysis based solely on observations finds much larger *SAD* values among weather stations located in different sub areas as revealed by the cluster analysis, than the *SAD* among five weather stations located in one sub area, i.e., inland continental Argentina. Additionally, the observations reveal very large *SAD* values among stations located in the river and Uruguay, a region in which the cluster analysis reveals more concentration of sub areas. In this sense it can be concluded that the analysis of observations unquestionably support the findings of the cluster analysis conducted with the BLM outputs.

Table 4. *SAD* values computed for the weather stations located in the river, Uruguay and two (AER and EZE) located in continental Argentina (inner rectangle of Figure 2, see text for details about PRE).

Site	PRE	COL	MGA	AER	EZE
PRE		24.8	45.0	32.0	15.9
COL			39.9	50.5	24.5
MGA				71.0	44.2

SAD (%) expresses the sum of the absolute values of differences between observed wind direction frequency roses (including calms).

5.8. Climatological aspects of spatial patterns

It is interesting to discuss the spatial pattern of the cluster distributions of Figure 4. Aside from the degree of detail provided by the increasing number of clusters, from 6 to 18, they all share a similar layout. The major aspects to be distinguished are that all cluster solutions tend to extend along the river, and that there is more change of cluster solutions towards the northeastern than the southeastern river shore. This disposition of the cluster solution is reflecting the main climatological features of the low-level wind field over the region, characterized by a sea-land breeze circulation already discussed in Section 2.

In order to assist in the interpretation of the cluster solution layout, an example is presented of the daily cycle of the low-level wind field in the region. The BLM is run for a typical summer day with northeasterly synoptic scale winds throughout the day, which is the dominant regional condition. Figure 6 presents the 10 m wind field calculated by the model at four different times of the day, from the morning until the evening. Throughout the day, most of the wind field changes take place over the river and neighbouring areas, in particular towards Uruguay where the wind directions change by more than one quadrant. This is because the afternoon inland component of the sea breeze tends to be in the opposite direction to the regional scale wind over the coast of Uruguay. Clearly, the regionality displayed by the cluster solution is in accordance with the spatial pattern of wind field changes throughout the day. Along coast the winds display a similar behaviour, which in turn depends on the distance to the coasts.

6. Conclusions

The outputs of a boundary layer model developed in previous reports for the La Plata River region of South America constitute the input for the present work. The model provides 180 17-dimension vectors that include 8-sector wind direction frequencies, frequency of calms, and mean wind speed of each wind sector. The four seasons of the year are studied but the analysis focalizes in summer that is the season that shows highest variability among the original individual vectors provided by the model and by the weather stations in the area.

In order to support further findings the model outputs for specific sites were contrasted against observations at nine weather stations belonging to Argentina and Uruguay. As a whole, this comparison showed good similarities taking into account the model error assessed in previous reports.

A first result of the hierarchical cluster analysis shows that the initial 180 vectors provided by the model can be well represented by a small number of them (cluster solution). In this sense, the advantage of the cluster analysis in synthesizing information becomes clearly evident when compared to the traditional method of visual inspection. The latter would require a detailed analysis of numerous grid point wind roses in order to achieve a similar result.

Three cluster solutions of 18, 12 and 6 groups are selected, as an example, to discuss their appropriateness to define regionality. Considering a sum for the absolute values of the

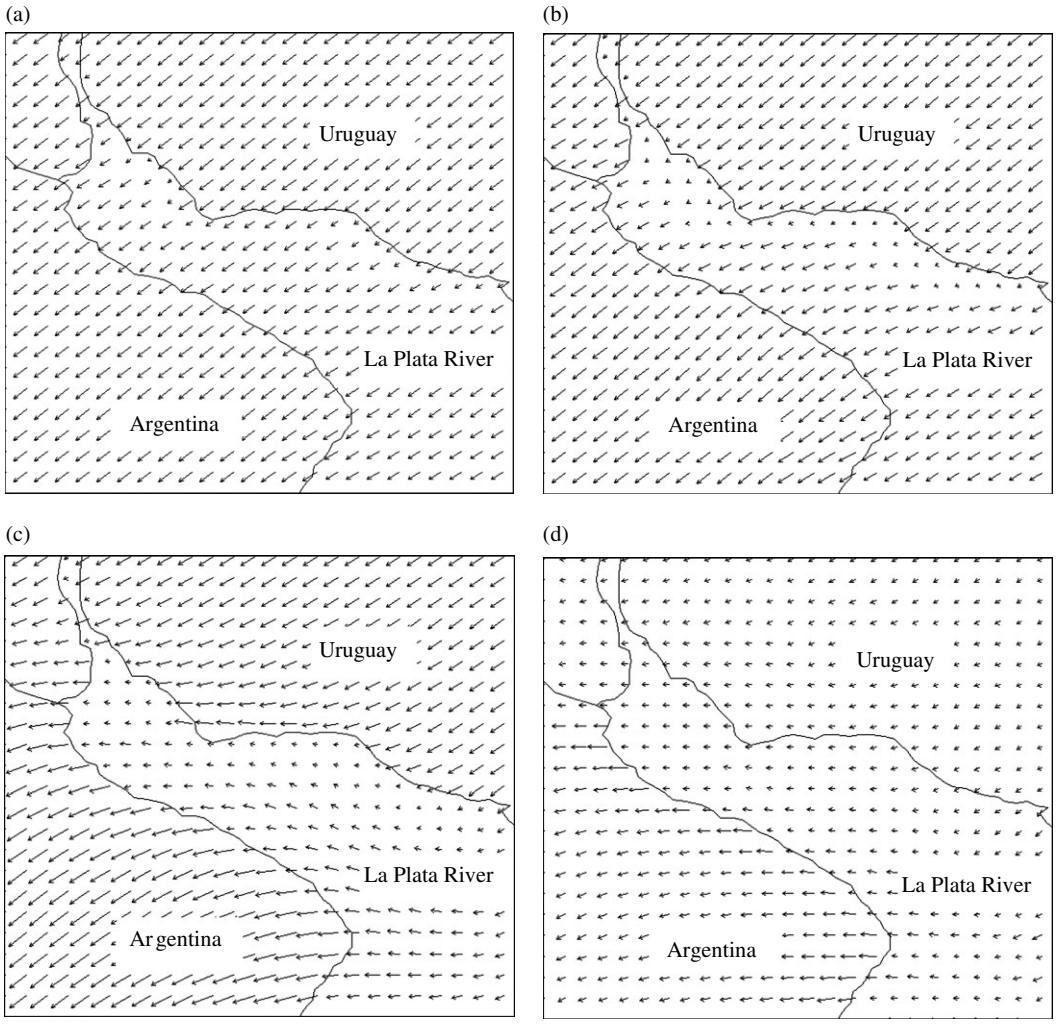

Figure 6. Example of 10 m wind field obtained with the BLM model at for a typical summer day, at (a) 1000 local time, (b) 1300 local time, (c) 1700 local time, and (d) 2100 local time. The wind scale is such that the maximum vector drawn is approximately 5 m s^{-1}.

differences (*SAD*) of 10% as a cut- off value for wind direction frequencies and calms it is possible to conclude that the 18-group cluster solution is somewhat redundant while the other two solutions are both quite appropriate with different degree of detail. The wind roses obtained as a result of the cluster solution provide the readers lacking information (Figure 5) about the behaviour of the wind fields in the sub areas of the area under study.

The analysis based solely on observations finds much larger *SAD* values among weather stations located in different sub areas as revealed by the cluster analysis, than the *SAD* among five weather stations located in one sub area, i.e., inland continental Argentina. Additionally, the observations reveal very large *SAD* values among stations located in the river and Uruguay, a region in which the cluster analysis reveals more concentration of sub areas. In this sense it can be concluded that the analysis of observations unquestionably support the findings of the cluster analysis conducted with the BLM outputs.

The actual distribution of weather stations in the region is not very far from the regionality that suggests the obtained cluster distribution. Thus, if new observing points were to be set up, the cluster solution can assist in the location of the new sites. The cluster distribution indicates that most of the wind field changes take place in a transect direction across the river rather than along coasts. Therefore, the new observing points would be more useful particularly over the river and towards the northern shore.

Acknowledgements

The authors are grateful to research grants PIP-0772 from Consejo Nacional de Investigaciones Cientŕficas y Técnicas and PICT2008-1417 from Agencia Nacional de Promoción Cientŕfica y Tecnológica of Argentina. The authors also acknowledge Servicio Meteorológico Nacional for providing the weather station data, with special thanks to Jose Ares. Besides, the authors thank the Editor for his helpful managing of the revision process.

References

Berri GJ, Sraibman L, Tanco R, Bertossa G. 2010. Low-level wind field climatology over the La Plata River region obtained with a mesoscale atmospheric boundary layer model forced with local weather observations. *J. Appl. Meteorol. Climatol.* **49**(6): 1293–1305.

Figueras S. 2001. *Análisis de Conglomerados o Cluster*. Universidad de Zaragoza: España.http://www.5campus.org/lección/cluster (accessed 17 March 2010).

Fovell RG, Fovell MC. 1993. Climate zones of the conterminous United States defined using cluster analysis. *J. Clim.* **6**: 2103–2135.

Gong X, Richman M. 1995. On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *J. Clim.* **8**: 897–931.

Kalstein LS, Tan G, Skindlov JA. 1987. An evaluation of three clustering procedures for use in synoptic climatological classification. *J. Clim. Appl. Meteorol.* **26**: 717–730.

Ratto G, Videla F, Maronna R, Flores A, De Pablo F. 2010a. Air pollutant transport analysis based on hourly winds in the city of La Plata and surroundings, Argentina. *Water Air Soil Pollut.* **208**: 243–257.

Ratto G, Maronna R, Berri G. 2010b. Analysis of wind roses using hierarchical cluster and multidimensional scaling analysis at La Plata, Argentina. *Boundary Layer Meteorol.* **137**: 477–492.

Romesburg C. 2004. *Cluster Analysis for Researchers*. Lulu Press: Morrisville, NC.

Unal Y, Kindap T, Karaka M. 2003. Redefining the climate zones of Turkey using cluster analysis. *Int. J. Climatol.* **23**: 1045–1055.

Wilks DS. 2006. *Statistical Methods in the Atmospheric Sciences*, 2nd edn. Elsevier: New York, NY.

Wolter K. 1987. The southern oscillation in surface circulation and climate over the Tropical Atlantic, Eastern Pacific and Indian Oceans as captured by cluster analysis. *J. Clim. Appl. Meteorol.* **26**: 540–558.

